DOCUMENT RESUME

ED 419 835 TM 028 410

AUTHOR Sarouphim, Ketty M.
TITLE Validity and Gender Issues in the Use of the DISCOVER Assessment for the Identification of Gifted Minorities.
PUB DATE 1998-04-00
NOTE 50p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS American Indians; Chi Square; *Construct Validity; Elementary Education; *Elementary School Students; Foreign Countries; *Gifted; Identification; Mexican Americans; *Minority Groups; Performance Based Assessment; *Sex Differences; Test Use

ABSTRACT

The purpose of this study was to investigate certain aspects of the internal structure of the DISCOVER assessment checklist (C. Maker, A. Nielson, and J. Rogers, 1994) to assess its construct validity. Other purposes included assessing percentages of identified students and gender differences through the use of the assessment. The sample consisted of 368 Navajo Indian and Mexican American participants from kindergarten, fourth, fifth, and sixth grades. The results showed low and nonsignificant inter-rating correlations, indicating high discriminant validity of the checklist. R-squared analyses revealed low percentages of variance explained, indicating low convergent validity. Items most frequently checked clustered around the underlying intelligence assessed. A pattern of more checks was found for higher ratings. Also, 22% of kindergarten, 27% of fourth- and fifth-grade, and 44% of sixth-grade participants were identified as gifted. Chi-square tests revealed no overall significant differences in the number of boys and girls identified as gifted. The results indicate sound qualities of the DISCOVER assessment and promote the use of performance-based assessments. An appendix contains checklist sample items. (Contains 8 tables and 38 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 419 835

Running Head: VALIDITY AND GENDER ISSUES

Validity and gender Issues in the Use of the DISCOVER Assessment for the Identification of Gifted Minorities

Ketty M. Sarouphim

Lebanese American University

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Ketty Sarouphim

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM028410

Abstract

The purpose of this study was to investigate certain aspects of the internal structure of the DISCOVER assessment checklist to assess its construct validity. Other purposes included assessing percentages of identified students and gender differences through the use of the assessment. The sample consisted of 368 Navajo Indian and Mexican-American participants from kindergarten, fourth, fifth, and sixth grades. The results showed low and non significant inter-rating correlations, indicating high discriminant validity of the checklist. R-squared analyses revealed low percentages of variance explained, indicating low convergent validity. Items most frequently checked clustered around the underlying intelligence assessed. A pattern of more checks was found for higher ratings. Also, 22% of kindergarten, 27% of fourth and fifth grades and 44% of sixth grade participants were identified as gifted. Chi-square tests revealed no overall significant differences in the number of boys and girls identified as gifted. The results indicate sound qualities of the DISCOVER assessment and promote the use of performance-based assessments.

Validity and gender Issues in the Use of the DISCOVER Assessment for the Identification of Gifted Minorities

The issue of identifying gifted students from culturally diverse groups has received much attention in the current literature (Baker, 1996; Maker, 1992; Clasen, Middleton, & Connell, 1994; Nielson, 1994; Scott, Perou, Hogan, & Gold, 1992). Several researchers have investigated the problem of overrepresentation of minority students in remedial programs and their underrepresentation in programs for the gifted (Clasen et al.; Gardner, 1992; Maker, 1993; Nielson, 1994). The often-cited causes for such practices are mostly traditional definitions of giftedness and narrow conceptions of intelligence as well as the use of traditional assessment procedures for identification purposes, such as standardized IQ tests (Clasen et al.; Cummins, 1991; Maker, 1992; Samuda, 1991). A constant source of dissatisfaction with standardized tests has been in the domain of fairness. Several studies on standardized tests have revealed gender, ethnic, and cultural bias ((Baker, 1996; Johnson, 1994). Sources of unfairness were attributed to the norms used for test interpretation, inadequacy of formats, bias in content, and linguistically loaded items (Baker, 1996). Consequently, other kinds of instruments are being considered and used, such as alternative assessments or performance-based assessments (clasen et al., 1994; Gardner, 1992; Maker, 1996).

Historically, giftedness was associated with superior

academic ability or achievement, measured by grade point average or IQ (Nevo, 1994). Terman's (1925) definition of gifted individuals as only those who scored in the top one percent in general intellectual ability on the Stanford-Binet Intelligence Test, exemplifies how giftedness was viewed more than a half century ago. Lately, however, a reconceptualization of the concept seems to have taken place, as evidenced by recent publications (Nevo, 1994). In 1972, a committee formed by the U.S. Office of Education (Marland, 1972) proposed a conception of giftedness that included not only abilities in the academic domain, but also in the performance domains. Areas in which children could be identified as gifted included high potential ability in (a) general intellectual ability, (b) specific academic aptitude, (c) creative or productive thinking, (d) leadership ability, (e) visual and performing arts, and (f) psychomotor ability.

Renzulli's (1979) three-ring definition of giftedness is another example of a broader view of the concept. He hypothesized that giftedness is an interaction between three clusters of basic traits, which are above-average general ability, high levels of creativity, and high levels of motivation (task commitment). Gifted individuals, according to Renzulli (1979), are those who possess this composite set of traits and are able to apply them to any area of human performance. Along the same lines, Maker (1993) postulated that creativity and intelligence were two components of the same construct. She contended that "creative

problem-solving" is a characteristic of giftedness. According to Maker (1996), the key element in giftedness is the ability to solve complex problems in the "most efficient, effective, or economical ways" (p. 44). Thus, gifted individuals are both highly intelligent and creative; their capabilities encompass not only understanding problems and discovering solutions using the most efficient methods, but also finding problems and solving them creatively and effectively (Maker, 1993, 1996).

In the same vein, the emergence of nontraditional theories of intelligence based on a broad conceptualization of intelligence have contributed to a reform of the concept as well. For example, Gardner (1983) defined intelligence as the multiple abilities that permit an individual to solve a problem or create a product that is valued within one or more cultural settings. In his book, Frames of Mind, Gardner (1983) rejected the unitary construct of intelligence and espoused a multidimensional definition in which he identified seven discrete intelligences: Linguistic, logical-mathematical, spatial, interpersonal, intrapersonal, bodily-kinesthetic, and musical. According to Gardner (1983), a person could be highly competent in one intelligence, but have average or low capability in the other intelligences. However, some intelligences are related, such as logical-mathematical and spatial abilities which Gardner (1983) called "twin" intelligences. Similarly, Sternberg (1991) proposed three kinds of intelligence: Analytic (the ability to understand the parts of a problem), synthetic (intuitive and creative

abilities) and practical (solving problems).

The new conceptions of giftedness and human intelligence warranted the development of different assessment techniques that extended beyond the use of standardized tests. However, the identification procedures in most states are still based on earning high scores on standardized tests (Kitano, 1991). Efforts have been made to introduce different measures, especially performance-based assessments (Clasen et al., 1994; Hafenstein & Tucker, 1994; Maker, 1996).

Proponents of performance assessment see many benefits associated with this technique (Frechtling, 1991). Among the advantages mentioned are assessing students in real lifelike situations, consideration of both process and product in evaluation, assessment of higher order skills, and use of more appealing material. Specific to the assessment of culturally diverse groups, the advantages often-cited include (a) nontraditional assessments usually are conducted in the dominant language of the person assessed and cover broad and multiple areas such as those advocated by Gardner (1983) and Sternberg (1991); (b) performance based assessments do not yield scores that will be transformed into standard z-scores to be compared with the scores of the normative sample; rather, evaluation of the individual is based on the judgment of multiple observers or evaluators, such as independent observers, parents and peers; even self nominations are considered valid (Baldwin, 1985); and (c) these methods are believed to be more fair and culturally

bias-free in comparison with multiple-choice questions that might require knowledge and skills specific to the dominant culture (Baldwin, 1985; Maker, 1992). However, alternative assessments are not without flaws. These new instruments have been criticized for their high cost, time-consuming procedures, domain underrepresentation, and mostly, their lack of sound psychometric qualities (Dunbar, Koretz, & Hoover, 1991; Frechtling, 1991; Hambleton & Murphy, 1992; Plucker, Callahan, & Tomchin, 1996). Because the scoring of performance assessments is complex and relies heavily on the judgment of observers or teachers, an element of subjectivity is introduced which renders determining the psychometric properties of these techniques a difficult task (Dunbar et al., 1991; Frechtling, 1991). Worthen (1993) contended that some evidence of the technical quality of alternative assessments is essential for this kind of tool to demonstrate its effectiveness and survive in the testing field.

Traditionally, sources of evidence for the validity of tests have centered mostly around the technical properties of the particular test under study. For example, Cronbach (1988) proposed the techniques of inspecting items, assessing internal correlations, and establishing correlation with other tests and practical criteria as procedures to assess construct validity. Similarly, for Messick (1995) sources of evidence include assessing convergent and divergent validity through the multi-trait multi-method matrix, examining the relationships among responses to items or parts of the test (internal structure), and

ruling out sources of invalidity which include construct underrepresentation, and construct-irrelevant variance. However, more recently, these researchers have added other non-technical elements to assess test validity, such as the social consequences of test use and interpretation, and administering the test to individuals who think aloud (Moss, 1992). According to Anastasi (1986) "almost any information gathered in the process of developing or using a test is relevant to its validity" (p. 3).

Different validation criteria have been proposed for performance-based assessments. Linn, Baker and Dunbar (1991) have argued that the traditional criteria used for test validation may not apply to performance assessment techniques. Instead, they proposed other criteria which include (a) fairness, (b) meaningfulness, (c) content coverage, (d) content quality, (e) generalizability, (f) consequences, and (g) cognitive complexity. A careful analysis of Linn et al. (1991) validation criteria for performance based assessment reveals a close resemblance to the criteria proposed by Messick (1995) and Cronbach (1988). The major difference is that Linn et al. (1991) have excluded the technical properties of the assessment as part of the validation criteria. However, Messick (1994) argued that "performance assessments must be evaluated by the same validity criteria... as are other assessments" (p. 13). Therefore, in this study, traditional quantitative criteria were used to assess the validity of a nontraditional performance-based assessment.

The use of performance-based assessments has been documented

in several studies. For example, Clasen et al. (1994,) conducted a well designed study in which they tested 433 minority and nonminority students using nontraditional multiple measures: problem solving, a free response drawing task, peer identification, and teacher nomination. The results showed that a total of 24% of the students tested were identified as gifted; minority and nonminority gifted students were identified in proportion to their actual distribution in the schools. Peer and teacher nominations supported the art and problem-solving identifications. Also, the number of males and females identified corresponded closely to their proportions in the population. The researchers concluded that nontraditional measures may be more culture and gender fair than are traditional assessments.

In another study, the effectiveness of a performance-based assessment in measuring intelligence was investigated (Hafenstein & Tucker, 1994). The assessment was designed according to developmentally appropriate tasks based on the theory of multiple intelligences and consisted of observing students perform tasks in different centers set to elicit behaviors relevant to the seven intelligences. Trained observers assessed three, four and five year old children as they rotated among the learning centers then classified their ability in the seven intelligences according to three rating categories: not evident, evident, and extremely evident. In mid-year teachers were asked to rate the children using the same criteria of the assessment rating scale. Content analysis of the two ratings revealed a great similarity

between observers' and teachers' rating of participants. Interviews with parents and teachers indicated that the use of the assessment led to adequate placement of the children. Regression analysis indicated that the beginning of year assessment was predictive of future performance. The researchers concluded that the performance-based assessment used in this study was an effective instrument in identifying young gifted children.

Using the conceptual framework of Gardner's theory of multiple intelligences and Maker's definition of giftedness (1993), Maker, Nielson, and Rogers (1994) have developed a performance-based assessment designed to identify gifted students among culturally diverse groups, called the DISCOVER assessment. The primary purpose of this study was to investigate certain aspects of the internal structure of the assessment checklist. Another purpose was to investigate the effectiveness of the DISCOVER assessment in reducing the problem of underrepresentation of minority students in programs for the gifted by determining whether through its use, a higher proportion of students from culturally diverse groups are being identified. A third purpose was to investigate gender differences in the number of identified students. Very few empirical studies on performance-based assessments have been reported in the literature (Baker et al., 1994; Plucker et al., 1996). Baker et al. (1994) stated that empirical studies on performance-based assessments constitute less than 5% of publications on the

subject. Considering the problem of minority underrepresentation in programs for the gifted and the high need for research on performance-based assessment, the significance of this study becomes evident. Four questions guided this study:

(1). How good is the fit between the DISCOVER assessment and Gardner's theory of multiple intelligences?

(2). What items of the checklist are most frequently checked by observers for students given the rating of "Definitely" (i.e., gifted) in each of the DISCOVER assessment activities? What items are least frequently checked across all four rating categories?

(3). What is the percentage of minority students identified as gifted through the use of the DISCOVER assessment?

(4). What proportions of males and females are identified through the use of the DISCOVER assessment?

Method

Participants

The sample of this study consisted of 368 students taken from kindergarten (n=114), fourth and fifth (n=141), and sixth (n=113) grades from six schools located in the northern and southern parts of Arizona. Participants were predominantly from culturally diverse groups (Navajo Indians and Mexican-Americans) and of low socioeconomic status as determined by place of residence and participation in the free lunch program. All participants were given the DISCOVER assessment in the fall of 1995 and the spring of 1996; some participants who were followed over a period of five years were administered the DISCOVER

assessment up to five times. Due to revisions in the DISCOVER assessment checklist in November 1995, only data collected in the 1996 spring administration of the assessment were used.

Instrument

The instrument used in this study is the checklist of the DISCOVER assessment. This performance-based assessment consists of five activities: Pablo®, Tangrams, Math, Storytelling, and Storywriting, designed to assess the problem-solving ability of students in spatial, logical-mathematical, and linguistic intelligences. Across the five activities, problems range from well-structured to unstructured situations devised according to the Maker-Schiever continuum of problem-types (Maker, 1993). A variation of tasks with increasing complexity are designed for aggregated grade levels (k-2, 3-5, 6-8, and 9-12).

In Pablo® (spatial), students are asked to make different constructions using colored cardboard pieces; in Tangrams (spatial/logical-mathematical), students start by making a geometrical shape using 21 Chinese tangrams pieces, then they solve puzzles using these pieces; in Math, worksheets consisting of problems increasing in difficulty and openness are used; in Storytelling (linguistic), students are provided with a set of toys and asked to either group or describe the toys depending on their grade level, then they are asked to tell a story of their choice, and in Storywriting (linguistic) students write a story of their choice. Interpersonal, intrapersonal and bodily-kinesthetic intelligences are not appraised through specific

activities, but behaviors pertaining to those intelligences are noted as students work in small groups throughout three of the five assessment activities (Pablo®, Tangrams, and Storytelling).

The length of a typical assessment session is approximately two and a half hours. Trained observers take notes and rotate at the completion of each activity to avoid observer bias. Following the assessment, observers meet to discuss the problem-solving behaviors of students and decide whether students are "Definitely", "Probably", or "Maybe" superior problem-solvers or whether their problem solving abilities are "Unknown" (Maker, 1996). The "Definitely" rating category is the highest and corresponds to giftedness. A student given the "Definitely" rating in at least two of the five DISCOVER assessment activities is identified as gifted. Observers complete a checklist for each child which incorporates the problem-solving behaviors of students (82 items) and the characteristics of their products (68 items) observed during each activity of the assessment. Observers check items corresponding to observed behaviors only once regardless of the intensity or duration of the occurrence. The checklist consists of items pertaining to each of the intelligences (except for musical intelligence which is not appraised in the DISCOVER assessment) and is divided into seven sections. (see Appendix A for sample items of the checklist). Each of the sections represents one intelligence with the last section entitled "General" consisting of items pertaining to the general traits of giftedness, such as full completion of tasks,

performance speed, and evidence of "problem-finding" behaviors. All of the DISCOVER checklist items represent superior problem-solving behaviors consistent with Gardner's (1983) description of "core capabilities" in each of the seven intelligences. (For more details on the assessment, see Maker, 1996).

Data Analysis

Participants' checklists were divided into three age groups according to the DISCOVER assessment variation of tasks across aggregated grade levels. Thus, three sets of identical, but separate statistical analyses were performed on the checklists of each of the three age groups (kindergarten, fourth and fifth, and sixth grade participants).

To determine the fit between the assessment and the theory of multiple intelligences, the convergent and divergent validity were calculated through a series of observers' inter-rating correlations. Observers' ratings of participants' problem-solving ability in all five activities were intercorrelated. Also, R-squared values were calculated for all significant correlations as a more accurate index of convergent validity.

To determine the items most frequently checked for students given the rating of "Definitely" in each activity, the frequencies and percentages of observers' checks for each item across the five activities were calculated. The operational definition of "items checked 50% or higher in the checklists of participants given the rating of "Definitely" across the three subsamples" was used to identify items that were most frequently

checked for the "Definitely" rating category.

To determine the percentage of participants identified and gender differences, a series of analyses were performed. First, the number of male and female participants who were given a rating of "Definitely" in each of the five activities was calculated. Then chi-square tests of significance for gender by activity were calculated to determine whether the difference in the number of male and female participants given a "Definitely" in each of the activities was statistically significant. The third step was to compute the number and percentage of male and female participants who were given a rating of "Definitely" in at least two of the DISCOVER assessment activities (i.e., identified as gifted). Then chi-square tests of significance for gender by gifted participants were performed to determine gender differences in each subsample.

Results

Fit Between the Assessment and the Theory of Multiple Intelligences

In Table 1, observers' inter-rating correlations are presented for the three subsamples and across all activities. As shown in Table 1, mostly low and non significant correlations were found in observers' inter-rating correlations for the kindergarten subsample; only one correlation was found significant and moderately high, the correlation between the variables of Storytelling and Storywriting (r_s (114) = 0.295, p < 0.01). In the fourth and fifth grade subsample, four correlations

were found significant and ranged from low to moderately high with the highest being between the variable of Storytelling and Storywriting ($r_s(141) = 0.35$, $p < 0.01$). In the sixth grade subsample, eight correlations were found significant and ranged from low to moderately high with the highest being between the variables of Math and Tangrams ($r_s(113) = 0.36$, $p < 0.01$).

R-squared or the percentage of variance explained was low for all significant correlations across the three subsamples (see Table 2) and ranged between 0.46 and 0.13 with the highest being between the variables of Math and Tangrams (13%) in the sixth grade subsample.

In sum, significant and moderately high correlations found between the variables of Storytelling (linguistic) and Storywriting (linguistic) in all three subsamples provide support for the convergent validity of the assessment. The same holds true for the significant and moderately high correlations between the variables of Math (logical-mathematical) and Tangrams (spatial/logical-mathematical) in the fourth and fifth, and sixth grade subsamples. However, the low R-squared values indicate low convergent validity of the checklist. On the other hand, the low and non significant correlations found between most of the other variables, especially in the kindergarten and fourth and fifth grades subsamples provide supportive evidence of the high divergent validity of the assessment.

Most and Least Frequently Checked Items

A pattern of higher percentages of item checks was found for

higher ratings, indicating that observers checked items at a higher frequency rate for participants given higher ratings. The increase in checks was gradual and in ascending order (i.e., students given the "Unknown" rating category received the least checks and students given the "Definitely" rating category received the highest frequency of checks) and consistent across all five activities.

Items most frequently checked for participants given the rating of "Definitely" were mostly items belonging to the respective intelligence assessed by the particular activities. A brief description of these items is found in Table 3. Also, a high percentage of checks were given to items belonging to the "General" section of the checklist. Interpersonal, intrapersonal, and bodily-kinesthetic items were the least checked across the three subsamples.

Items that received zero frequency checks across the three subsamples were either items that represent behaviors and characteristics of products which pertain to intelligences unrelated to the specific activity performed or items describing rare occurrences. For example, in Tangrams, items that were the least frequently checked were linguistic items and in Storytelling and Storywriting items with least frequency checks were either spatial items or items that describe the use of more than one language. (See Table 4 for a brief description of items least frequently checked).

Percentage of Identified Participants

In Table 5, a distribution of observers' ratings of the participants' problem-solving ability is presented. Data are displayed according to each subsample and across the five activities. Because the two activities of Math and Storywriting are performed a day preceding or following the three other activities, missing data represent absenteeism on either date.

As shown in Table 5, the "Unknown" rating category had the lowest frequencies in all activities and across grade levels, with the exception of the Storytelling activity in kindergarten. The highest frequency of observers' ratings fluctuated between the "Maybe" and "Probably" categories, depending on the activity and grade level. Also, the sixth grade participants had the highest percentages of "Definitely" ratings whereas the kindergarten participants had the lowest percentages of "Definitely" ratings, except for Storywriting in which that pattern was reversed. Moreover, in Math, kindergarten participants were found to have a higher percentage of "Definitely" ratings than fourth and fifth grade participants. On the other hand, observers' ratings showed the lowest variation among sixth graders (i.e., observers' ratings of participants' problem-solving ability in the five activities were more clustered around higher ratings) whereas the ratings were less evenly distributed among kindergarten and fourth and fifth grade participants.

Gender Differences

Gender by activity.

As shown in Table 6, the number of boys given the rating of "Definitely" in Pablo® exceeded that of girls in all grade levels. However, the difference was not found statistically significant (see Table 7), except for the sixth grade participants (χ^2 (1, n = 113) = 4.622 , $p < 0.05$).

In Tangrams, an equal number of kindergarten boys and girls were given the rating of "Definitely"; that number was slightly higher (less than one percent) for girls in sixth grade and moderately, but not significantly higher for girls (9.2%) in the fourth and fifth grades. No significant gender differences were found in Tangrams for all grade levels, as indicated by the low and non significant chi-square values (see Table 7).

In Math, the number of kindergarten boys given the rating of "Definitely" exceeded that of girls by about one percent whereas the number of girls given the rating of "Definitely" exceeded that of boys by 2.8% in fourth and fifth grade participants and by 4.4% in sixth grade participants. No significant gender differences were found in Math for all grade levels, as indicated by the low and non significant chi-square values (see Table 7).

In Storytelling, the number of girls given the rating of "Definitely" exceeded that of boys in all grade levels. The difference was not significant in sixth graders (less than one percent) nor in kindergarten participants (less than two percent). However, a statistically significant difference was found in the fourth and fifth grade subsample, with the number of girls given the rating of "Definitely" exceeding that of boys by

11.4% (χ^2 (1, n = 141) = 5.443, p < 0.05).

In Storywriting, more girls than boys (2.6%) were given the rating of "Definitely" in kindergarten. The same was found for the other two subsamples with the number of girls given the rating of "Definitely" exceeding that of boys by 5.6% in fourth and fifth graders and by 6.2% in sixth grade participants. However, none of the differences was found statistically significant, as indicated by the low and non significant chi-square values (see Table 7).

Gender by gifted participants.

As indicated in Table 8, 22.8% of kindergarten participants were identified as gifted, that is given the rating of "Definitely" in at least two of the DISCOVER assessment activities. A slightly higher percentage was found for fourth and fifth graders (27.7%) whereas the percentage of sixth graders identified as gifted was about twice that of kindergarten participants (44.2%).

In terms of gender differences, a statistically equal proportion of boys and girls were identified as gifted in all three subsamples. In kindergarten, an equal number of boys and girls were identified as gifted; a slightly higher percentage of girls were identified as gifted in sixth grade (1.8%) and a moderately, but not significantly higher percentage of girls were identified as gifted in fourth and fifth grades (7.9%). In terms of statistical significance, the difference between the number of boys and girls identified as gifted was not found significant in

all grade levels, as indicated by the low and non significant chi-square values (see Table 8). In other words, through the use of the DISCOVER assessment, no gender differences were found in the identification of gifted participants in kindergarten, fourth and fifth, and sixth grade subsamples.

Discussion

In this study, some of the psychometric properties of the DISCOVER assessment checklist were investigated through a series of statistical analyses pertaining to the internal structure of the checklist. The percentages of identified students and gender differences were also assessed. High discriminant validity, but relatively low convergent validity was found for the checklists of the three subsamples as suggested by the low inter-rating correlations. In assessing convergent and discriminant validity, an index of reliability needs to be calculated for further evidence of true variance (Anastasi, 1988). Griffiths (1996) found high interobserver reliability in the two studies she conducted on the Pablo® activity of the DISCOVER assessment. In both studies, percentages of agreement between observers ranged between 80 and 100%, and differences in observers' levels of experience did not affect significantly the percentage of agreements. Therefore, the variance in the ratings of the present study reflect mostly true variance since a major source of error variance in the ratings (observers' drift) was not found significant, as suggested by the high interobserver reliability (0.81, $p < 0.01$) in Griffiths' (1996) studies.

In terms of the number of identified students and gender differences, the results of this study showed that large percentages of participants were identified across the three subsamples with the highest being among sixth graders. Also, chi-square tests revealed statistically significant differences in Pablo® in favor of boys among the sixth grade subsample and in Storytelling, in favor of girls among fourth and fifth graders. However, no overall statistically significant differences were found in the numbers of males and females identified as gifted across the three subsamples.

The findings of this study suggest a good fit between the assessment and the theory of multiple intelligences. In other words, separate intelligences are appraised through each of the activities of the DISCOVER assessment, as shown by the sufficiently low and non significant correlations found between most of the variables. Thus, participants found gifted in one intelligence were not necessarily found gifted in the other intelligences. On the other hand, correlations between observers' ratings of students' ability in activities in which similar or related intelligences are assessed were found significant, suggesting that these activities, namely Storytelling and Storywriting (linguistic) and Math and Tangrams (spatial\logical-mathematical) appraise similar abilities. These findings are congruent with the results of other studies. For example, Benbow and Minor (1990) found that linguistic and mathematical intelligences were two distinct abilities, as indicated by the

significant differences in the scores of mathematically and linguistically gifted individuals, suggesting that mathematical and linguistic abilities represent two separate intelligences. Also, in the same study, a significantly greater number of boys was found spatially gifted whereas more girls were identified as verbally gifted. Similar results were obtained by Plucker et al. (1996) who found that a significantly higher number of girls were identified as linguistically gifted through the use of nontraditional measures, but no significant differences were found in the overall number of boys and girls identified as gifted.

An unexpected result is the large number of significant inter-rating correlations found among sixth graders. One interpretation may be related to the particular procedures followed with sixth graders over the last five years. Some of the sixth grade participants had been given the DISCOVER assessment as many as five times whereas for some others, the spring 1996 assessment (data used in this study) was their first experience with the DISCOVER assessment. This could explain why sixth graders' ratings were more clustered around similar ratings than participants in the other two subsamples. Separate analyses could not be performed to isolate the practice effect because of the small sample size of sixth grade participants who were assessed only once.

Also, a finding of significance is that the majority of items identified as characterizing the "Definitely" category in each of

the activities pertain to the respective intelligence assessed. This is meaningful because it shows that the checklist items represent one clustered entity integrated within the intelligence measured. Another explanation may be that the items are more focused on specific rather than general abilities, a finding compatible with the theory of multiple intelligences.

The identified items most frequently checked for participants given the "Definitely" rating category corroborate the findings of a study conducted to identify universal identifiers of giftedness (Coleman, 1994). In this study, teachers were trained to identify gifted students using portfolios. A committee was formed to search the literature for a list of universal identifiers. The committee decided on four categories; these are: Exceptional learner, exceptional user of knowledge, exceptional generator of knowledge, and exceptional motivation. In Coleman's article, the identifiers established as evidence for giftedness are highly similar to the items identified in this study as most frequently checked for students given the "Definitely" rating category. For example, the identifiers "shows high levels of inquiry and reflection" and "learns quickly and easily" (Coleman, 1994, p. 66) resemble to a great extent the DISCOVER checklist items "exhibits problem-finding ability" and "solves problems quickly". In both, the abilities of using higher order skills and speed of operation are highlighted in different ways.

Another important finding is related to the high percentage

of checks given to some of the items in the General section of the checklist. At first sight, this finding seems to undermine the good fit between the theory of multiple intelligences and the DISCOVER assessment. However, a careful scrutiny warrants a different interpretation. Maker's conceptualization of giftedness is broad and encompasses a wide variety of abilities subsuming intelligence, creativity, giftedness and talent (Maker, 1993). According to Leung (1981), giftedness is composed of specific as well as absolute traits. Absolute traits are general and found in all gifted individuals regardless of their highly developed ability in a specific domain of intelligence. Even though the DISCOVER assessment is based on the theory of multiple intelligences with which a good fit was found, a general component as well was included in the assessment. The General section of the checklist reflects Maker's belief in the absolute traits of giftedness, whereas the other sections reflect her conceptualization of giftedness as composed of specific traits. The finding of a high percentage of checks given to the checklist General items across activities and grade levels provides support to the absolute nature of giftedness. The finding that the majority of the items identified as characterizing the "Definitely" category in each of the activities pertain to the specific intelligence appraised gives support to the specific nature of giftedness. Hence, both general and specific traits of giftedness are appraised through the DISCOVER assessment.

Moreover, the findings of this study revealed that most

items pertaining to interpersonal, intrapersonal, and bodily-kinesthetic intelligences did not receive a large percentage of checks in any of the activities and across grade levels. Whereas this finding provides support for the validity of the checklist (i.e., items mostly checked in each of the activities pertained to the corresponding intelligence measured), it also shows that interpersonal, intrapersonal and bodily-kinesthetic intelligences are not measured comprehensively in the DISCOVER assessment. Therefore, specific activities need to be developed with grade appropriate tasks for a more accurate domain representation and appraisal of these intelligences.

Another interesting finding is the relatively high percentage of participants identified as gifted. This finding is congruent with the results of other studies in which a performance-based assessment was used as the instrument for identification. For example, in the study conducted by Clasen et al. (1994), the final pool of identified students included 24% of the participants. In another study, Hafenstein and Tucker (1994) found that the use of a performance assessment based on the theory of multiple intelligences was an effective procedure to identify young gifted students. One possible explanation for the relatively large percentage of identified participants in this study may be the grounded theory on which the DISCOVER assessment is based. Given the nature of multiple intelligences, the possibility of identifying gifted minority participants using the DISCOVER assessment is higher than that in traditional

assessments in which a full scale IQ normed mostly on the majority population is used for identification procedures. Adherents of a full scale IQ claim that gifted individuals are those with extremely high scores (two or two and a half standard deviations above the mean), thus constituting three to five percent of the population. Hence, in their view, giftedness is unidimensional and of one kind only. However, if we embrace the view advanced in the theory of multiple intelligences, giftedness takes many forms and becomes of a multidimensional nature. Statistically, the probability of identifying gifted students through the use of the DISCOVER assessment is much higher than that found in traditional tests of intelligence. By definition, through the use of the DISCOVER assessment, an individual is identified as gifted if he or she is given the rating of "Definitely" in at least two of the activities. Given that the DISCOVER assessment is composed of five activities, each individual could be identified as gifted through ten different combinations (i.e., Pablo® and Tangrams, Pablo® and Math, Pablo® and Storytelling, Pablo® and Storywriting, Tangrams and Math, Tangrams and Storytelling, Tangrams and Storywriting, Math and Storytelling, Math and Storywriting, Storytelling and Storywriting). Thus, through the use of the DISCOVER assessment, the probability of identifying giftedness in the population is largely increased which might explain the high percentage of participants identified as gifted across grade levels in this study.

The results of this study provided evidence in support of the construct validity of the DISCOVER assessment checklist. A good fit between the assessment and the theory of multiple intelligences, in addition to a well-balanced internal structure suggest sound psychometric properties of the checklist. Moreover, given the historically ineffective assessment of minorities and their underrepresentation in programs for the gifted, this study has shown that the use of the DISCOVER assessment with culturally diverse groups may reduce the problem of minority underrepresentation in programs for gifted students. Also, the absence of gender differences provided additional evidence of the fairness of the DISCOVER assessment. However, only a limited profile of the checklist technical aspects was drawn. In future research, other aspects need to be investigated, such as the internal consistency of the checklist, the effect of observers' gender and years of experience on their rating of students, and pre and post reliability studies. Other limitations of this study include the sample composition of students from only two culturally diverse groups (Mexican-Americans and Navajo Indians) and a restricted range of grade representation. In further research, samples need to include students from other culturally diverse groups (e.g., Asians, African-Americans) and from upper grade levels to support the use of the DISCOVER assessment with populations of different ethnicities and ages.

In this study, the findings have shown that the psychometric qualities of performance based assessments may be examined using

traditional quantitative measures and still meet the criteria for valid assessments. The sound technical properties of the DISCOVER assessment found in this study along with the other advantages associated with performance-based assessments (e.g., real lifelike situations, assessments of higher order skills, consideration of both process and product, lack of cultural bias) might lead to a more frequent use of nontraditional assessments. Perhaps future use of performance based assessment along with monitored studies may demonstrate empirically the need for a paradigm shift in the testing field.

References

Anastasi, A. (1986). Evolving concepts of test validation. Annual Review of Psychology, 37, 1-15.

Anastasi, A. (1988). Psychological testing. (6th ed.). New York: Macmillan Publishing Company.

Baker, E. L. (1992). The role of domain specification in improving the technical quality of performance assessments, project 2.2: Alternative approaches to measuring liberal arts subjects. Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA. (ERIC Document Reproduction Services No. 346 133).

Baker, E. L. (1996). Introduction to theme issue in educational assessment. Journal of Educational Research, 89(4), 194-196.

Baker, E. L., O'Neil, H. F., & Linn, R. L, (1994). Policy and validity prospects for performance-based assessment. Journal for the Education of the Gifted, 17(4), 332-353.

Baldwin, A.Y. (1985). Programs for the gifted and talented: Issues concerning minority populations. In F. Horrowitz & M. O'Brien (Eds.), The Gifted and Talented: Developmental Perspectives (pp. 223-249). Washington, D.C.: American Psychological Association.

Benbow, C. P., & Minor, L. L. (1990). Cognitive profiles of verbally and mathematically precocious students: Implications for identification of the gifted. Gifted Child Quarterly, 34, 21-26.

Clasen, D. R., Middleton, J. A, & Connell, T. J. (1994).

Assessing artistic and problem-solving performance in minority and nonminority students using a nontraditional multidimensional approach. Gifted Child Quarterly, 38(1), 27-37.

Coleman, L. (1994). Portfolio assessment: A key to identifying hidden talents and empowering teachers of young children. Gifted Child Quarterly, 38(2), 65-69.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer (Ed.), Test validity (pp. 3-17). Hillside, NJ: Erlbaum.

Cummins, J. (1991). Institutionalized racism and the assessment of minority children: A comparison of policies and programs in the United States and Canada. In R.J. Samuda, S.L. Kong, J. Cummins, J. Pascual-Leone, & J. Lewis. (Eds.), Assessment and placement of minority students (pp. 97-107). Kingston/Toronto: Intercultural Social Sciences Publications.

Dunbar, S. B., Koretz, D. M., & Hoover, H. D. (1991). Quality control in the development and use of performance assessments. Applied Measurement in Education, 4, 289-303.

Frechtling, J. A. (1991). Performance assessment: Moonstruck or the real thing? Educational Measurement: Issues and Practices, 10(4), 23-25.

Gardner, H. (1983). Frames of mind: The theory of multiple intelligences. New York: Basic Books.

Gardner, H. (1992). Assessment in context: The alternative to standardized testing. In B. R. Gifford, & M. C. O'Connor. (Eds.), Changing assessments: Alternative views of aptitude,

achievement, and instruction (pp. 77-120). Boston: Kluwer.

Griffiths, S. E. (1996). The inter-observer reliability of the DISCOVER problem-solving assessment. Unpublished manuscript, University of Arizona at Tucson.

Hafenstein, N. L., & Tucker, B. (1994). Performance-based assessment: An alternative assessment process for young gifted children. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Serves No. 370 994).

Hambelton, R. K., & Murphy, E. (1992). A psychometric perspective on authentic measurement. Applied Measurement in Education, 5, 1-16.

Johnson, N. E. (1994). Use of the WISC-R with disadvantaged gifted children: Current practice, limitations and ethical concerns. (ERIC Document Reproduction Services No. ED 368 097)

Kitano, M. (1991). A multicultural educational perspective on serving the culturally diverse and non-diverse gifted students. Journal for the Education of the Gifted, 15, 4-19.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: expectations and validation criteria. Educational Researcher, 20(8), 5-12.

Leung, E.K. (1981). The identification and social problems of gifted bilingual-bicultural children. Paper presented at the Council for Exceptional Children. (ERIC Document Reproduction Services No. ED 203 653)

Maker, C. J. (1992). Intelligence and creativity in Multiple

Intelligences: Identification and development. Educating Able Learners, XVII(4), 12-19.

Maker, C. J. (1993). Creativity, intelligence, and problem solving: A definition and design for cross-cultural research and measurement related to giftedness. Gifted Education International,9, 68-77.

Maker, C. J. (1996). Identification of gifted minority students: A national problem, needed changes and a promising solution. Gifted Child Quarterly, 40, 41-50.

Maker, C. J., Nielson, A.B., & Rogers, J.A. (1994). Giftedness, diversity, and problem-solving. Teaching Exceptional Children, 27(1), 4-19.

Marland, S. P., Jr. (1972). Education of the gifted and talented (V.I.). Report to the Congress of the United States. Washington, DC: U.S. Government Printing Office.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessment. Educational Researcher, 23(2), 13-23.

Messick, S. (1995). Validation of inferences form person's responses and performance as scientific inquiry into score meaning. American Psychologist, 50, 741-749.

Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. Review of Educational Research, 62(3), 229-258.

Nevo, B. (1994). Definitions, ideologies, and hypotheses in gifted education. Gifted Child Quarterly, 38, 184-186.

Nielson, A. B. (1994). Traditional identification: Elitist, racist, sexist? New evidence. Communicator: The Journal of the California Association for the Gifted, 24(3), 18-31.

Plucker, J. A., Callahan, C. M., & Tomchin, E. M. (1996). Wherefore art thou, multiple intelligences? Alternative assessments for identifying talent in ethnically diverse and low income students. Gifted Child Quarterly, 40, 81-90.

Renzulli, J. S. (1979). What makes giftedness? Re-examining a definition. Phi Delta Kappan, 60, 180-184.

Samuda, R. J. (1991). Psychometric factors in the appraisal of intelligence. In R. J. Samuda, S. L. Kong, J. Cummins, J. Pascual-Leone, & J. Lewis. (Eds.), Assessment and placement of minority students (pp. 25-40). Kingston/Toronto: Intercultural Social Sciences Publications.

Scott, M. S., Perou, R., Urbano, R., Hogan, A., & Gold, S. (1992). The identification of giftedness: A comparison of white, Hispanic and black families. Gifted Child Quarterly, 36, 131-139.

Sternberg, R. J. (1991). Giftedness according to the triarchic theory of human intelligence. In N. Colangelo, & G.A. Davis (Eds.), Handbook of Gifted Education (pp. 45-54). Boston: Allyn & Bacon.

Terman, L. (1925). Mental and physical traits of a thousand gifted children. In L. Terman (Ed.). Genetic studies of genius (Vol 1). Stanford: Stanford University Press.

Worthen, B. R. (1993). Critical issues that will determine the future of alternative assessment. Phi Delta Kappan, 74, 444-454.

Table 1

Observers' Inter-Rating Correlations for Kindergarten, Fourth and Fifth Graders, and Sixth Grade Participants

Activity	1	2	3	4	5
			Kindergarten (n = 114)		
1. Pablo®	--	0.067	0.128	0.117	0.094
2. Tangrams		--	0.172	0.162	0.147
3. Math			--	0.021	0.190
4. Story				--	0.295*
5. Writing					--
			Fourth and Fifth (n = 141)		
1. Pablo®	--	0.123	0.083	0.053	0.121
2. Tangrams		--	0.331**	0.257**	0.135
3. Math			--	0.215*	0.174
4. Story				--	0.354**
5. Writing					--
			Sixth (n = 113)		
1. Pablo®	--	0.300**	0.235**	0.227*	0.165
2. Tangrams		--	0.361**	0.307**	0.113
3. Math			--	0.305**	0.218*
4. Story				--	0.218*
5. Writing					--

*p < 0.05. **p < 0.01.

Table 2

R-Squared and Percentage of Variance Explained for Statistically Significant Inter-Rating Correlations in all Three Subsamples

	Kindergarten		4th & 5th		Sixth Grade	
Correlations	R^2	%	R^2	%	R^2	%
Pablo® / Tangrams	--	--	--	--	0.09	9.0
Pablo® / Math	--	--	--	--	0.055	5.5
Pablo® / Story	--	--	--	--	0.051	5.1
Tangrams / Math	--	--	0.109	10.9	0.130	13.0
Tangrams / Storytelling	--	--	0.066	6.6	0.094	9.4
Math / Storytelling	--	--	0.046	4.6	0.093	9.3
Math / Storywriting	--	--	--	--	0.047	4.7
Storytelling / Storywriting	0.087	8.7	0.125	12.5	0.047	4.7

Table 3

Description of Most Frequently Checked Items for the "Definitely" Rating Category

Pablo®
Spatial: Constructions are complex, symmetrical, detailed, three-dimensional, and show attention to design. Student makes various types of constructions.
Intrapersonal: Student shows pleasure when problem is solved.
General: Student is involved, enjoys and completes tasks, and shows increased motivation for open-ended problems.
Tangrams
Spatial: Student takes pieces off, does not manipulate pieces before fitting them, and is first in the group to complete puzzles.
Logical-Mathematical: Student solves puzzles in logical and various ways.
General: Student is involved, enjoys and completes tasks, is continuously working and persists on difficult problems.
Math
Logical-Mathematical: Student solves problems correctly, creates and solves many problems, and has higher scores than classmates.
General: student completes all tasks.
Storytelling
Linguistic: Student is fluent and gives details. Stories have a plot, action, logical and appropriate sequence of events.

Table 3 (continued).

Description of Most Frequently Checked Items for the "Definitely" Rating Category

Storytelling (Continued)
Interpersonal: Stories demonstrate understanding of other people's emotions, motivations, and of social relationships.
General: Student completes tasks and is involved and continuously working.
Storywriting
Linguistic: Student is fluent. Stories have a plot, action, and appropriate sequence of events.
General: Student completes tasks.

Table 4

Description of Items with Zero Frequency Checks

Pablo®
Bodily-kinesthetic: Handwriting shows good motor coordination.
Tangrams
Linguistic: visual images are created for listener.
Spatial: Constructions show movement.
Bodily-kinesthetic: Student supplements story with body or toys. Handwriting shows good motor coordination.
General: Products reflect an environment other than that in which student evolves.
Math
Logical-mathematical: Student creates patterns and uses mathematical concepts in products.
General: Products are unique.
Storytelling - Storywriting
Linguistic: Student uses more than one language to tell story.
Logical-mathematical: Student creates patterns.
General: Products are unique.

Table 5

Frequencies and Percentages of Observers' Ratings of Participants' Problem-Solving Ability in each Activity Across Grade Levels

Rating	Pablo®		Tangrams		Math		Story		Writing	
	n	%	n	%	n	%	n	%	n	%
Kindergarten										
Unknown	9	7.9	16	14.0	14	12.3	34	29.8	8	7.0
Maybe	45	39.5	63	55.3	39	34.2	37	32.5	38	33.3
Probably	38	33.3	23	20.2	36	31.6	27	23.7	33	29.0
Definite	22	19.3	12	10.5	25	21.9	16	14.0	23	20.2
Missing	0	0.0	0	0.0	0	0.0	0	0.0	12	10.5
Total	114	100	114	100	114	100	114	100	114	100
Fourth and Fifth										
Unknown	4	2.8	13	9.2	13	9.2	15	10.6	14	10.0
Maybe	43	30.5	39	27.7	44	31.3	50	35.5	36	25.5
Probably	66	46.8	39	27.7	43	30.5	48	34.1	46	32.6
Definite	27	19.2	49	34.7	26	18.4	26	18.4	22	15.6
Missing	1	0.7	1	0.7	15	10.6	2	1.4	23	16.3
Total	141	100	141	100	141	100	141	100	141	100

Table 5 (continued).

Rating	Pablo®		Tangrams		Math		Story		Writing	
	n	%	n	%	n	%	n	%	n	%
				Sixth Grade						
Unknown	2	1.8	11	9.7	15	13.2	7	6.3	8	7.1
Maybe	35	31.0	33	29.2	34	30.1	37	32.7	40	35.4
Probably	37	32.7	26	23.0	29	25.7	27	23.9	34	30.1
Definite	37	32.7	41	36.3	27	23.9	39	34.5	21	18.6
Missing	2	1.8	2	1.8	8	7.1	3	2.6	10	8.8
Total	113	100	113	100	113	100	113	100	113	100

Table 6

Frequencies and Percentages of Boys and Girls Participants Given the Rating of "Definitely" in each Activity Across Grade Levels

	Frequency			Percentage		
Grade	Boys	Girls	All	Boys	Girls	All
			Pablo®			
Kindergarten	14	8	22	12.3	7.0	19.3
4th & 5th	15	12	27	10.6	8.5	19.1
Sixth	23	14	37	20.3	12.4	32.7
			Tangrams			
Kindergarten	6	6	12	5.3	5.3	10.6
4th & 5th	18	31	49	12.8	22.0	34.8
Sixth	21	20	41	18.6	17.7	36.3
			Math			
Kindergarten	13	12	25	11.4	10.5	21.9
4th & 5th	11	15	26	7.8	10.6	18.4
Sixth	11	16	27	9.7	14.1	23.8

Table 6 (continued).

Grade	Frequency			Percentage		
	Boys	Girls	All	Boys	Girls	All
			Storytelling			
Kindergarten	7	9	16	6.1	7.9	14.0
4th & 5th	5	21	26	3.5	14.9	18.4
Sixth	19	20	39	16.8	17.7	34.5
			Storywriting			
Kindergarten	10	13	23	8.8	11.4	20.2
4th & 5th	7	15	22	5.0	10.6	15.6
Sixth	7	14	21	6.2	12.4	18.6

Table 7

Chi-Square Tests of Significance for Gender by Activity Across Grade Levels

Activity	Kindergarten		Fourth & Fifth		Sixth	
	df	Chi-square	df	Chi-square	df	Chi-square
Pablo®	1	2.586	1	2.100	1	4.622*
Tangrams	1	0.017	1	0.516	1	0.314
Math	1	0.181	1	0.002	1	0.513
Story	1	0.151	1	5.443*	1	0.007
Writing	1	0.153	1	0.901	1	2.291

*$P < 0.05$

Table 8

Chi-square Tests of Significance for Gender by Gifted Participants Across Grade Levels

Grade	Boys		Girls		All		df	Chi-square
	n	%	n	%	n	%		
Kindergarten	13	11.4	13	11.4	26	22.8	1	0.042
4th & 5th	14	9.9	25	17.8	39	27.7	1	0.454
Sixth	24	21.2	26	23.0	50	44.2	1	0.016

Appendix A

Checklist Sample Items

PROBLEM-SOLVING BEHAVIORS

C. June Maker

Judith A. Rogers

Aleene B. Nielson

The University of Arizona

Name_____School _____
Date_____Grade _____ Birthdate _____
_____ M F Teacher _____Obs# ____
___Language_____Ethnicity _____
_____Observer/Checklist:_____Pablo __
_____ Tangrams _____ Story
Observers:_____Math_____
_____Writing

By Activity

1.Pablo:Strength? _____unknown _____maybe_____probably_____
_definitely

2.Tangrams:Strength? _____unknown_____maybe_____probably_
_____definitely

3.Math:Strength? _____unknown _____maybe_____probably___
____definitely

4.Story:Strength? _____unknown_____maybe_____probably____
___definitely

5.Writing:Strength? _____unknown_____maybe_____probably__
_____definitely

Student Name _____

1.Linguistic: Strength? _____ unknown_____ maybe _____ probably _____definitely

1.1 Problem-Solving	Pablo	Tang.	Math	Story	Writ.
1.1.1 tells stories easily and fluently					
1.1.2 uses more than one language					
1.1.3 chooses colorful or unusual adjectives and adverbs					
1.1.4 invents and plays with words					
1.1.5 gives descriptions easily and fluently					
1.1.6 translates concepts from one language to another					
1.1.7 changes voice to represent different characters					
1.1.8 other _____					
1.1.9 other_____					

1.2 Product(s)	Pablo	Tang.	Math	Story	Writ.
1.2.1 descriptions are detailed					
1.2.2 stories have recognizable beginning, middle, and end					
1.2.3 stories/pictures have a recognizable plot					
1.2.4 stories/pictures have action					
1.2.5 vocabulary includes complex and/or sophisticated words or concepts					
1.2.6 stories/pictures include emotions and feelings					
1.2.7 stories/pictures include humor					
1.2.8 invented spelling shows understanding of word sounds/meanings					
1.2.9 stories/pictures include dialogue or conversation					
1.2.10 visual images are created for the reader or listener					
1.2.11 stories include comparisons of similarities or differences					
1.2.12 stories include complex sentences or syntax					
1.2.13 stories/pictures include complex ideas, (e.g., philosophical, moral, spiritual, political, cultural)					
1.2.14 stories include all pieces given during activity					
1.2.15 titles of stories/pictures correlate with content of story					
1.2.16 stories/pictures have a sequence of events that is appropriate to the story					
1.2.17 stories/pictures include cause-effect relationships					
1.2.18 stories/pictures have a sense of the macabre					
1.2.19 stories/pictures create an emotional response in the reader					

BEST COPY AVAILABLE